DOCUMENT RESUME

ED 194 510 SP 016 997

AUTHOR Powell, J. C.
TITLE Developing a Training Program to Improve Teacher Effectiveness. A Prospectus for Research Proposals.
SPONS AGENCY International Business Machines Corp., Armonk, N.Y.: Ontario Educational Research Council, Toronto.
PUB DATE 80
NOTE 39p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Academic Achievement: *Cognitive Processes: *Diagnostic Teaching: Elementary Secondary Education: Expectation: Inservice Teacher Education: Interaction: Learning Processes: *Multiple Choice Tests: Problem Solving: Teacher Behavior: Teacher Effectiveness: *Teaching Methods: Test Construction

ABSTRACT

The selection of wrong answers for multiple choice tests appears not to be blind guessing, as the current theory assumes, but represents systematic choice based upon the ways in which learners extract meaning and solve problems. It appears the departures from "expected" answers arise when the learner interprets the question in a manner different from that intended, uses a solution strategy differently than intended (perhaps incorrectly), or works from different information (perhaps misinformation) to solve problems. It would appear that the learner is unaware of these differences from expectation, and gives the best answer that he or she is able. The limiting factor confounding educators would seem to be inclinations to expound expectations rather than exploring those of the learners. Both teachers and learners may be unaware that their meaning extraction and thinking processes are not equivalent. To overcome this problem an alternative teaching strategy and testing procedure which is derived from wrong answer information is proposed, and a description is given of a teacher inservice course based on these procedures. (JD)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED194510

DEVELOPING A TRAINING PROGRAM TO IMPROVE TEACHER EFFECTIVENESS

A

PROSPECTUS

FOR RESEARCH PROPOSALS

by

J. C. POWELL

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY
J. C. Powell

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

© 1980 J. C. Powell
Portions of the projects in this report have been funded by research grants from International Business Machines (IBM) and from the Ontario Educational Research Council (OERC)

SP 016 997

ABSTRACT

This report gives a fairly detailed overview of a considerable amount of research which is related to the educational use of wrong answer information. It draws a parallel between the work in the area of interpersonal relationships done by Argyris and Schön and the work of the present author in the area of cognition. In both cases, tacit assumptions, related to the habitual ways in which we derive meaning from our environment appears to constrain our effectiveness as coworkers and/or learners. Since we are unaware of these assumptions, and therefore, their impact, and since the typical method of teaching does not expose them for examination, we have been unaware that our actions have been placing an unnecessarily low ceiling upon learning achievement. Measurement experts have not assisted us in the detection of this problem because they have, typically, replaced wrong answer information with "zeros" before analyzing tests and/or test results. This action has inadvertently removed these important variables from further scrutiny. An alternative teaching strategy and testing procedure which is derived from wrong answer information is proposed. The procedures are also illustrated in the first appendix. The technical details of the observations which have thus far been made concerning the distributions of wrong answers are summarized in a second appendix. An annotated bibliography is given in the third appendix.

DEVELOPING A TRAINING PROGRAM TO IMPROVE TEACHER EFFECTIVENESS

INTRODUCTION

A number of years ago I encountered a situation where a teacher asked her class to draw "seven blue robin's eggs" in a nest. One child drew only five eggs because he "knew" that robins never lay more than five eggs. How should this teacher mark this assignment? What is a "WRONG ANSWER?"

In this presentation I will address a single hypothesis. I will assume that participants in the educative process are capable of learning substantially more than they actually accomplish. Second, I will assume that there is something about the current approach to education which constrains its effectiveness. Third, I will hypothesize that this constraining property is related to the way in which the day-to-day operations of classroom interactions are conducted. Specifically, I will hypothesize that the way in which teachers and learners currently extract MEANING from their interactions is the limiting factor.

To test this hypothesis, I propose to develop an in-service training program designed to alter the procedures of a group of teachers in how they approach meanings. I will try to determine, by longitudinal follow-up, whether or not substantially more learning is achieved when this alternative approach is used.

I will begin by explaining why I chose this particular hypothesis; discussing my reasoning in three directions: historical, psychological, and statistical. This prospectus will conclude with a brief explanation of the hypothesis in terms of teaching practice, a brief description of the current status of the proposed in-service program, and an outlook toward future research. An appendix details the general features of the proposed program.

HISTORICAL BACKGROUND

Except that programmed instruction seems to help learning occur more rapidly than most other approaches, the general finding about the relative effectiveness of different methods of teaching has been largely inconclusive, suggesting no fundamental advantage to any procedure.

These null results are both discouraging and intuitively unsatisfying. They are discouraging because, if true, they imply that we may have reached the maximum of human educability. They are unsatisfying because the intuition of classroom experience suggests that different approaches should make a difference to learning.

Individual experiences seem to contradict these research results. On certain occasions, progress in "learning" substantially beyond "normal expectations" has been reported. Maria Montessori, for instance, tutored a group of "retarded" children to a point where they passed the Italian Elementary School Examinations.

If schooling has actually reached the upper limits of human capacity, such events would never occur. A more reasonable explanation would seem to be that <u>something</u> about the way we teach may be putting an unnecessarily low ceiling upon current achievement. This conclusion supports the two assumptions which form the background to this study, but leaves unanswered the question concerning what this "something" might be. Also, this "something" must be an aspect of the way we teach about which we are unaware; otherwise our attempts to change educational effectiveness would have been more successful.

My personal experience has led me to a particular conclusion about what this "something" might be.

In my undergraduate training as a teacher I encountered a book about the teaching of singing to elementary school boys, which was essentially a diagnostic-prescriptive approach, which

worked on the premise: "This is how you recognize this particular problem, and here is what to do about it." I found this approach very exciting and was grossly disappointed to find my training in the teaching of mathematics, the physical sciences, and in the language arts (the so-called CORE SUBJECTS) did not follow this same format.

We talked a bit about the importance of Self-Concept to learning effectiveness in psychology but we did not develop strategies for helping learners to improve their self-concepts. Of course, this was the middle 1950's and much has been learned in 25 years.

Shortly after I started to teach (secondary school, math/sciences) I took a course in how teachers should construct tests. The procedure we used for item analysis involved making large analysis tables by hand. (This was before computers were readily available). After arranging student responses in order of total-correct scores, we were to use an "X" for a response if it were "wrong" and a "blank" for a response if it were "right."

In my tabulation I made a fortuitous modification of this approach by using a code for the particular "error" for wrong answers. In most cases I used multiple choice tests and I merely entered the letter of the actual incorrect choice.

Once this modification was being used, I began to notice that particular "wrong" answers seemed to cluster in rather narrow performance ranges on the total-correct score scale. Closer examination of these particular errors showed me the particular types of mistakes certain students were making.

In effect, the wrong answers contained <u>diagnostic</u> information.

I began to use this information in my teaching and my students began to perform better on the schoolwide tests we set in our department. That is, <u>where I looked</u> for learner information made a difference in <u>what I could do</u> to help them to learn.

As I worked with this information, I began to notice that

wrong answers on several items seemed to overlap. But this overlap was not always the product of the same "error type." At this point, to solve the puzzle, I began to enquire how the students were attempting to solve the problems which showed "errors."

To my surprise, errors which seemed different in my logical analysis were actually very similar in that they were the product of the same general solution procedure.

For instance, students in algebra who used substitution of simple whole numbers as their strategy of first choice to get the answers but who were unsure of or careless about the signs of the numbers and/or order of operations could make all sorts of seemingly unrelated errors and still get apparently difficult questions correct.

Thus it became evident to me that how the students approached their math and science problems had an important effect upon their success on tests.

The issue was not, as the music book suggested, "which error" but rather "why any error at this point." The way a student attacked a problem seemed to be directly related to both, a) which questions he or she got right, and, b) to the particular errors made when the answer was wrong.

Once I started to teach these students from the points of view they were using, a surprising thing happened. Both their enthusiasm and their performance made a large jump upward, the latter being measured on our school-wide tests. Both this enthusiasm and improvement resulted from the fact that this procedure helps the learners to look at the reasoning processes which are producing errors.

They tended, as a rule, to average somewhat more than a half standard deviation above the average for the balance of the school. Voluntary remedial students gained and sustained an average gain of 30 marks. The more I departed from the traditional approaches,

making sure to use their view-points, and to measure carefully as I proceeded, the better they did. My greatest achievement for average students in this context was in my eighth year of teaching when my two classes (out of seven in the same subject in the school) were a full standard deviation above the balance of the school. A randomized approach to initial class placement and a common year-end test jointly marked was used, leaving little doubt about the validity of this anomalous event.

This is not the end of the story. My success with aversge learners led me to being asked to teach below average performers. My best results in Secondary School was an overall average increase in standardized reading score performance from a Grade Equivalent of 3.6 in October to Grade Equivalent of 7.2 in the following February. When I tried this approach with City Center Elementary School children (Grade 6) their reading score advance was 6 times as large as it had been the previous year ($\triangle$1.8 as compared with $\triangle$ 0.3). When I tried this approach as a school psychologist working with other teachers one child gained 20 IQ points on the Standard Binet. This is the only child for whom we had highly reliable pre- post- data. Many others seemed to show similar performance gains.

By this time, I was convinced that the children we teach in school appeared to be capable of learning much more than they commonly achieve, if we focus upon the reasoning and understanding lapses which cause their errors. On the other hand, my secondary school special class youth had difficulty (as did Kohl's 36 children) dealing with the LABEL "special" when they attempted the "regular" program for which their performance gains would seem to have equipped them. Two years in the special program made the transition easier than did just one year; and making the procedures I was using explicit (i.e. helping them to think about their own thinking) seemed to assist the adjustment more than keeping this information to myself. (Bateson's Deutero learning).

It might be argued that what I was doing was merely good teaching. Clearly, I was more knowledgeable about testing than most practicing teachers. In this case, mandatory courses in test construction for teacher trainees should be the answer. However, such courses, when mandated, do not seem to lead to measureable improvement in learner performance. I must have been doing something <u>different</u> from many other teachers in order to have had this impact. The question was, "What was this difference?" Both <u>where I looked</u> and <u>how I used</u> the resulting data seemed to be involved.

THE PSYCHOLOGICAL ASPECTS

I attempted to identify this difference in my Doctoral Dissertation and my subsequent research.

The first discoveries I made were in the Qualifying Project for my Dissertation. I used a language-based rather than a mathematical test. I asked the examiners to write an explanation for each answer selection. I then used their reported reasoning as a guide for interpreting the statistically homogeneous groups of "wrong" answers obtained from factor analysis. I used the most common reasoning reported by one group to classify the error types and with a second group to predict their reasoning on the same wrong answer selections. About two-thirds of the reported reasons of the second group were equivalent to the most common reasoning for the first group. This was rather a surprising finding for answers which were supposed to be "guesses."

Even more surprising was the observation that among the other third of the reasons given were explanations for answers which in strictly logical terms were quite correct. It seemed incongruent if not paradoxical that "wrong" answers could be correct, but nonetheless, this was the observation.

Even more fascinating were the insights these reports of reasoning gave me into the way in which these students attacked these problems.

In a subsequent study using the same test, I carried this approach even further, trying to tap the developmental sequences in reported reasoning.

A single example will illustrate the psychological findings of this latter study.

FIGURE 1

EXAMPLE INTERPRETATION

PROVERB:

Quickly come, quickly Go.
(Easy Come, Easy Go)*
Translations:

A. ALWAYS COMING AND GOING AND NEVER SATISFIED.
Characteristic of 13 year olds.
"You Should Stick To A Job Til It's Finished."

B. WHAT YOU GET EASILY DOES NOT MEAN MUCH TO YOU.
Characteristic of adults.

C. ALWAYS DO THINGS ON TIME.
Characteristic of 8 year olds.
"That's What Teacher Always Says."

D. MOST PEOPLE DO AS THEY PLEASE AND GO AS THEY PLEASE.
Characteristic of 10 year olds.
"It Talks About Coming and Going."

* Item No. 18 from Gorham, Donald R. PROVERBS TEST
Psychological Test Specialists, 1956.
Reproduced with permission.

In Figure 1 I have presented a four-alternative multiple choice test item from the test I am using for my studies.

For the 8-year olds, their most common interpretation of "QUICKLY COME, QUICKLY GO" was "Always do things on time."

The most common explanation for this choice (more than half of those who chose this answer gave this reason) was "That's what the teacher always says." I assume these children have interpreted QUICKLY COME, QUICKLY GO in terms of physical movement; "I QUICKLY COME" and "I QUICKLY GO." In these statements combined we have an accurate description of a closely time-tabled Grade 3 classroom. Finishing work in the allotted time is indeed imperative. The answer is perfectly correct. The frame of reference is physical action rather than metaphorical-ascription.

This particular "wrong" answer, rather than being a "blind guess" seems to be a product of the "theory of meaning" being used by the learner. In this case the theory of meaning seems to be "What I DO is meaningful to ME." The discrepancy between their answers and the "right" ones arises from differences in the "theory of meaning" in use.

These children have interpreted this question quite correctly, in a manner not intended by adults.

In the case of the most common answer for the ten-year olds, QUICKLY COME, QUICKLY GO is taken to mean "Most people do as they please and go as they please." From their typical reasoning, "It talks about coming and going," I infer that their theory-of-meaning-in-use relates to a word-for-word translation of the statement into other words. Their theory-of-meaning-in-use seems to be to search for literal meanings such as dictionary definitions.

Closer examination of these particular errors suggested to me that the particular types of mistakes these students were making are now "detached" or "dissociated" from their personal experience. In Piaget's terms, they have "decentered."

A similar interpretation can be made for the typical answer for the 13 year olds. In this latter case the cause-effect relationship implied in the "never satisfied" aspect of their choice seems to be the basis for their answers. Thus one step further in the development toward formal logic has been taken.

The "correct" answer is a judgmental statement about behavioral expectations which is ascribed by our culture. It is a metaphore whose meaning is generally accepted but which has no necessary validity.

As we can see, then, choice changes among wrong answers in this particular case describes a developmental sequence with "meaning" becoming progressively more remote from directly observable events as it also becomes more "logical."

This pattern has proven to be very strong developmentally and largely typical rather than atypical under cross-validation and replication conditions.

An important second conclusion seems to be that the wrong (or unexpected) answers which learners give seem to tap their "theories of meaning in use."

More recently, I have encountered the Argyris and Schön concept of Single-Loop and Double-Loop learning. This approach is remarkable to me in that they observed that people behave in accordance with a "Theory-in-use" which is related to meaning extraction very similar to my findings. They were working on interpersonal issues where I was working on cognition.

The second property they observed seems to be the critical issue. People generally have made a habit of their theory-in-use and are, therefore, unaware of the limitations it puts upon their behavioral effectiveness. In both cases, (cognitive and interpersonal behavior) these two properties seem to be invariant (i.e. they are characteristic of nearly everyone). I give an example of this property for cognition in the appendix.

In summary, then, the selection of wrong answers for multiple choice tests appears <u>not</u> to be blind guessing, as the current theory assumes, but represents systematic choice based upon the ways in which learners extract meaning and solve problems. It appears that departures from "expected" answers arise when the learner interprets the question in a manner different from that

intended, used a solution strategy differently than intended (perhaps incorrectly) or works from different information (perhaps misinformation) to solve the problem. It would appear that the learner is unaware of these differences from expectation, and gives the best answer he or she is able.

From a psychological basis, the limiting factor confounding educators would seem to be our current inclination to expound our expectations rather than exploring those of the learners with them. Both teachers and learners therefore may be unaware that their meaning extraction and thinking process are not equivalent. Perhaps neither understands why the other is failing to communicate as expected. Thus, schools appear to fail to educate learners who depart in any marked degree from the patterns of expectation held by the teachers. Schools may also be expected to limit the learning outcomes of those in the "dominant culture" to the constraints in viewpoint held tacitly by this culture.

STATISTICAL ASPECTS

I shall not go into considerable technical detail with respect to my statistical findings since these are already elsewhere in the literature. For those interested I have provided a technical note. Instead, I will look briefly at a number of tacit assumptions which result from "classical test theory," and indicate what my findings reveal.

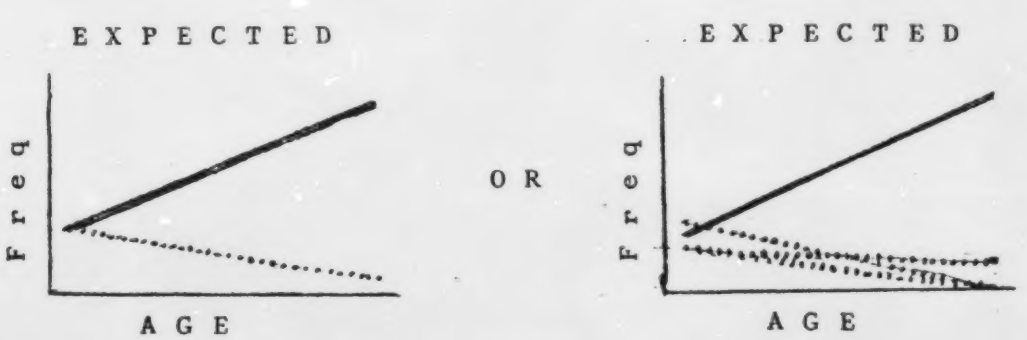
If we assume that the learner either "knows or guesses" there should be no meaningful statistical patterns among wrong answer selections. Specifically, we would expect to observe the following outcomes:

1. Frequencies among the selection of different wrong alternatives will be about equal to each other whatever the age level.
2. Frequencies of wrong answer patterns will appear as nearly parallel, if not equivalent straight lines of downward slope with age.

3. Wrong answers will not predict the results of independent measures of learner success.
4. If the impact of frequencies is removed from the data there should remain no meaningful relationships among pairs of wrong alternatives (other than random statistical artifacts).
5. The only way a learner will show progress from one administration of a test to another is to change from any wrong answer to the right answer on any particular question.

My observations show that each one of these expectations is false.

First, the pattern of frequence distribution for wrong answers seems to be some form of peaked function, with different alternatives within an item having their peak frequencies at different age levels. Wrong alternatives are expected to look like this:



but actually look more like this;

As we can see from these diagrams, both of the first two expectations fail to appear in the observed distributions.

Second, if the total correct scores on one test are predic-

ted by statistical homogeneous combinations of right and wrong answers from another test, then the wrong answers are the best predictors (using multiple regression analysis). Also, curved line analysis gives better predictions than does straight line analysis. In this finding the third expectation is disconfirmed.

Third, if the impact of frequencies are removed, then relationships among pairs of wrong answers occur with much greater frequency than among pairs of right answers. These occur much more frequently than chance, tend to occur in the same direction (joint choice rather than mutually exclusive choice), and change in pattern and frequency with age. In fact, the relative importance of wrong answer pairs, both on terms of frequency and predictive power appears to increase with age. In these findings, the fourth expectation is disconfirmed.

Finally, there is strong evidence to suggest that learner progress is shown by changing from a particular wrong answer to the right answer rather than any wrong answer, from a particular wrong answer to another specific wrong answer, and, in some cases, from the right answer to a particular wrong answer. The data analysis for this latter conclusion is not yet complete, but should be for a report to the NCME in April 1980.

Although data analysis is far from complete, it is already very clear that the pattern of answer selection is FAR MORE meaningful than the number of right answers.

How have researchers missed these observations before this? By the simple expedient of treating all wrong answers as a "zero" (0) at the time of scoring the test and in so doing (except for using a "guessing correction") dropping them from further analysis. The result of this procedure has been to make these five hypotheses (and several others) into tacit assumptions and to make us unaware that they were false by removing from possible study the data (wrong answers) we needed to disconfirm them.

In conclusion, once again we have a set of hypotheses rend-

ered untestable by our problem solving procedures, binding us to the limitations of the "dominant culture" in our society as defined in this context by the "right" answers on our tests.

It is now necessary to link this discussion to typical classroom practice in order to suggest how the making of such undetected errors in the schooling process might have led to a self-limiting set of learning outcomes.

RELATIONSHIP TO TEACHING

The usual teaching procedure is to establish a set of expectations (behavioral objectives) for the learners before we teach, and to interact with the class members in ways which induce them to produce the responses we are expecting. We then test them to determine the degree to which they can reproduce these same expectations. Essentially we LISTEN FOR our own expectations. This approach tends to be at arm's length from the directly observable data in our classroom environment.

There is an alternative approach. We ask a question, and then explore with the learners how they arrive at the answers they give us. Instead of focusing upon the answers we focus our attention upon how the questions were devised and how the answers to these questions were formulated. Instead of <u>listening for</u> an expected answer we LISTEN TO the learner. We pay close attention to the actual learning (or meaning generating) events.

In each of these two approaches the psychological environment is being manipulated but the <u>manner</u> of this manipulation is different.

In the LISTENING FOR approach the linkage <u>between</u> the self-limiting tacit assumptions of the learners and the desired outcomes of the teachers is tacitly assumed rather than being explicitly established. For this reason the teacher can easily be <u>misled</u> into assuming that children who answer as expected <u>understand</u> what they have been "taught." It is one thing to recognize

a concept as an insightful resolution of some data problems with which we are intimately familiar. It is quite another to be told a conclusion without an understanding of the data problems on which it is based.

The robin's eggs illustration at the beginning of this paper is a good example. The boy had an intimate knowledge about robins. The teacher was concerned about the abstraction of "sevenness" and making it "concrete."

The child who cannot frame his or her ideas the way the teacher does very quickly learns to be a failure.

Thus, the accumulation of nonunderstanding and/or misunderstanding over several years would be expected to provide an upper limit to the amount of effective learning which can be achieved.

In the LISTENING TO approach the linkage between how each child is thinking and where the teacher desires that each learner go is explicit and public from the outset and remains so throughout the process.

Those answers which are valid within their own context are immediately recognizable as such. In this situation the child need never be bewildered why his or her RIGHT answer was called "wrong." When a learner is WRONG the exploration of his or her reasoning reveals why the error occurred. Detection and correction of error becomes built into the life expectations of each learner. As each gains skills in this latter ability they may be expected to become self-instructional.

Within this alternative, the focus within the classroom does not center upon the collected wisdom of our culture, but rather how this wisdom was generated. The content of subject matter in the classroom need not change - but the method of using it most certainly would change.

Learner status estimates would shift from summative and formative approaches to some form of response pattern analysis. What each child can do would be more important than how much he or she can recite. It should be noted that "criterion referenced"

approaches contain the same implicit error of focussing only upon "right" answers, inherent to other current approaches to educational evaluation.

Examples of these procedures from the three areas of language arts, physics, and mathematics are given in Appendix A. The first of these will explore an "elaborative" teaching procedure related to the exposure of tacit assumptions. In the physics example an illustration of how wrong alternatives on a multiple choice test can be designed to be diagnostic is provided. A discussion of how to use this information in the LISTENING TO mode is then presented. Finally, in the mathematics example, I focus primarily on several different ways to solve a typical High School algebra problem. Having learners explore alternative solution strategies is another method for elaboration.

CONCLUSION

We began our discussion with the suggested hypothesis that incompatabilities in the way in which teachers and learners extract MEANING from their interactions, may be a major limiting factor to the educative process.

It is now possible to clarify this hypothesis. The tacit assumption made by teachers that the child who can recite a statement understands it, combined with the tacit assumption of the learner that if he or she can recite a statement he or she has met the teacher's expectations, coupled with the tacit assumption during measurement that wrong answers indicate a lack of learning, in concert, serve to constrain the effectiveness of the educative process. Furthermore, the tacitness of these underlying assumptions have served to make us unaware of the cause of this unnecessarily low ceiling to learning.

This rather involved statement should be directly testable. If the teacher stops making this tacit assumption and, instead, helps the learners explore their own ways of generating meaning along with the ways proposed by others; if the teachers use the

nature of and basis for unexpected answers as important data in the instruction process, then these hypothesized limiting processes should be reduced.

With this limitation removed the depth of understanding should increase and the learners should progress more rapidly than is "normal." By using what currently exists as "normal," it should be relatively easy to confirm or refute this hypothesis.

There may be other limiting factors which are also unknown but which are independent of this one. This exploration of learner behavior the LISTENING TO approach requires, might expose these factors as well.

Thus, this study will be confirmatory in the sense that a major improvement in the effectiveness of instruction would identify at least one of the operational factors currently reducing instructional effectiveness.

To achieve the test of the central hypothesis in this proposed study, a program of four full (at least 60 hour) courses of in-service teacher education is proposed. One course will involve specific training in the design of observational techniques (including classroom tests) such that all the learners' answers are useful to the teacher. This course is now in modularized form with one successful field trial. Extensive revision is in progress.

The second course, which will focus upon the personalization of instruction within the group context is designed to develop the skills needed to use the information generated in the "Testing" course. It is currently in a rather loose modularized form and has had several trial runs.

Both of these courses will be taught in the first year of the project. The other two courses will be taught one each in the second and third years of the project. Their content will be derived from the classroom experiences of the teachers in the project and from technical considerations related to the topics approved for funding. A three year follow-up is also suggested.

The scale, duration, target population and other details of

this project will be worked out in accordance with the specific interests of granting agencies.

The remarkable similarities between the findings of Argyris and Schön in the interpersonal field and these findings in the cognitive field would seem to lend mutual support to both positions. That others often limit their thinking by the way in which they attack a problem has long been known. That we, ourselves, are subject to these same limitations and that we are UNAWARE of these limitations is currency of a much more recent vintage.

Now that a means of exposing and exploring these overlooked errors by establishing a Model O-II (double loop or "exploratory") learning system has been discovered and is under development, a substantial forward step in human capability may be in the offing. In this latter case, the outlook for further research would be vast. Virtually all educational research here-to-fore conducted would probably need to be reworked within the framework of this alternative context. The study proposed herewith could provide a beginning for this prospect.

APPENDIX A

This appendix will illustrate the critical elements which will be included in the program to be used in this project. Three illustrations will be given, one from each of: Language Arts, Physics, and Mathematics. In each case a different aspect of the program will be illustrated in detail and followed by a brief discussion of how this example can be extended into more complex learnings.

Language Arts Example:

Task: Combine these two sentences into a single one in four different ways.

1. John hit the ball.
2. The ball went through the window.

The resulting product is collected and examined for the presence or absence of sentences of the type:

John hit the ball AFTER it went through the window.

Once this last sentence has either been collected (a rare event) or failing that (which is much more common) presented, a brief time is spent collecting possible scenerios for this latter sentence and other various possibilities.

In a separate place, objections which are raised to this sentence and to the scenerios are, also, collected (on the chalk board, for instance).

The proponents of the scenerios are then asked to indicate the assumption which they made to produce their scenerios. Those who object are also asked to indicate their assumptions.

Typical assumptions made BEFORE the exploration include:

1. The "ball" in each of the two sentences is the SAME ball for both events.
2. The time sequence of these events is the same as the order of the sentences.
3. John is the causal agent of the second event.

Secondary assumptions often include:

1. The ball went from outside to inside and not the reverse.
2. The window broke.
3. Etc.

Once these assumptions have been collected and shown to be not necessarily true, the instructor then advocates the following three hypotheses to be explored with the group.

1. The tacit assumption made about these two sentences determined how they were "combined" and the limits of the range of production witnessed in this demonstration.
2. The larger the number of tacit assumptions made before "combination" the more difficult it is for a person to entertain alternative "combinations" outside of the initial set of relationships.
3. Once these tacit assumptions have been exposed the range of possibilities for "combination" is expanded greatly.

The evidence collected in this exercise should be sufficient to test these hypotheses. The generalization from these conclusions can itself be presented as a testable hypothesis:

General Hypothesis

The degree that we fail to surface and enquire into the logical validity of the assumptions we make about a problem is directly related to the degree to which our personal thinking ability is limited:

An interesting observation from this exercise is that people are unaware of how much they may be limiting their own creative capabilities until the manner and nature of these limitations is exposed to them. The exercise illustrates three important points.

First: Teachers rarely consider the assumptions a learner has to make before answering a question nor the impact of

these assumptions upon the answers given, and, therefore, their ability to learn.

(The Listening for approach)

Second: When a teacher shifts to an "exploratory approach" (one more conducive to the implementation of a Model 0 - II learning system) then "creativity" is enhanced, and so is motivation (i. e. the Listening to approach).

Third: The ability of a person to consider alternative possibilities in a problem situation, to cast it into other frames of reference, and in general to manipulate his or her environment symbolically before taking physical action is teachable by the procedure illustrated by this exercise.

Whole ranges of possibilities may be opened in this manner. It appears to be of critical importance in this approach that we begin with the assumptions actually made by the learners themselves. By using this approach, cultural differences, learning style differences, inconsistencies of viewpoint from differing information sources etc., all become publicly confrontable. The way in which we generate knowledge about our world rather than the knowledge others have generated becomes a subject matter of education.

In this alternative approach the accumulated wisdom of mankind still remains the content of instruction, but no longer the sole focus for teaching nor the only product observed when assessing learner status and learner progress.

This shift of orientation from "content" to "process" seems to work best when it is accompanied by using the same approach to interpersonal and other forms of "action" learning (in the terms of Argyris and Schön when a Model 0 - II learning system has been established). Perhaps this observation may arise from the more basic observation that the tacit underlying assumptions people make which limit their thinking seem to be consistent and

seem to be related to their "meaning extraction" strategies.

Part of the teacher's ability in this respect can be assisted by the design of observation procedures (such as tests) which tap the way in which the learners are viewing the problem. It is this issue which we will consider next.

Physical Science Example

The research reported in the prospectus has shown there are three general ways in which "errors" are generally either produced or selected. These are:

1. Using a frame of reference different from the one intended by the questioner.
2. Using a different procedure (or using a procedure incorrectly) than the one intended by the questioner.
3. Using a different information base (including misinformation) than the one intended by the questioner.

Each of these three:

Frame, Procedure, and Information Base if themselves correct will produce unexpected BUT CORRECT answers. Answers, then, actually appear to be a complex interaction among these three operational factors; how the learner extracts MEANING, what the learner DOES, and what the learner already KNOWS.

The design of the test items to tap these aspects of learner status is somewhat more complex than the procedure advocated by test designers. In multiple choice items, for instance, the advocacy is for distractors to be "plausible."

The example I give here illustrates this more complex approach. In this case this particular item isolates the KNOW and DO dimensions as follows:

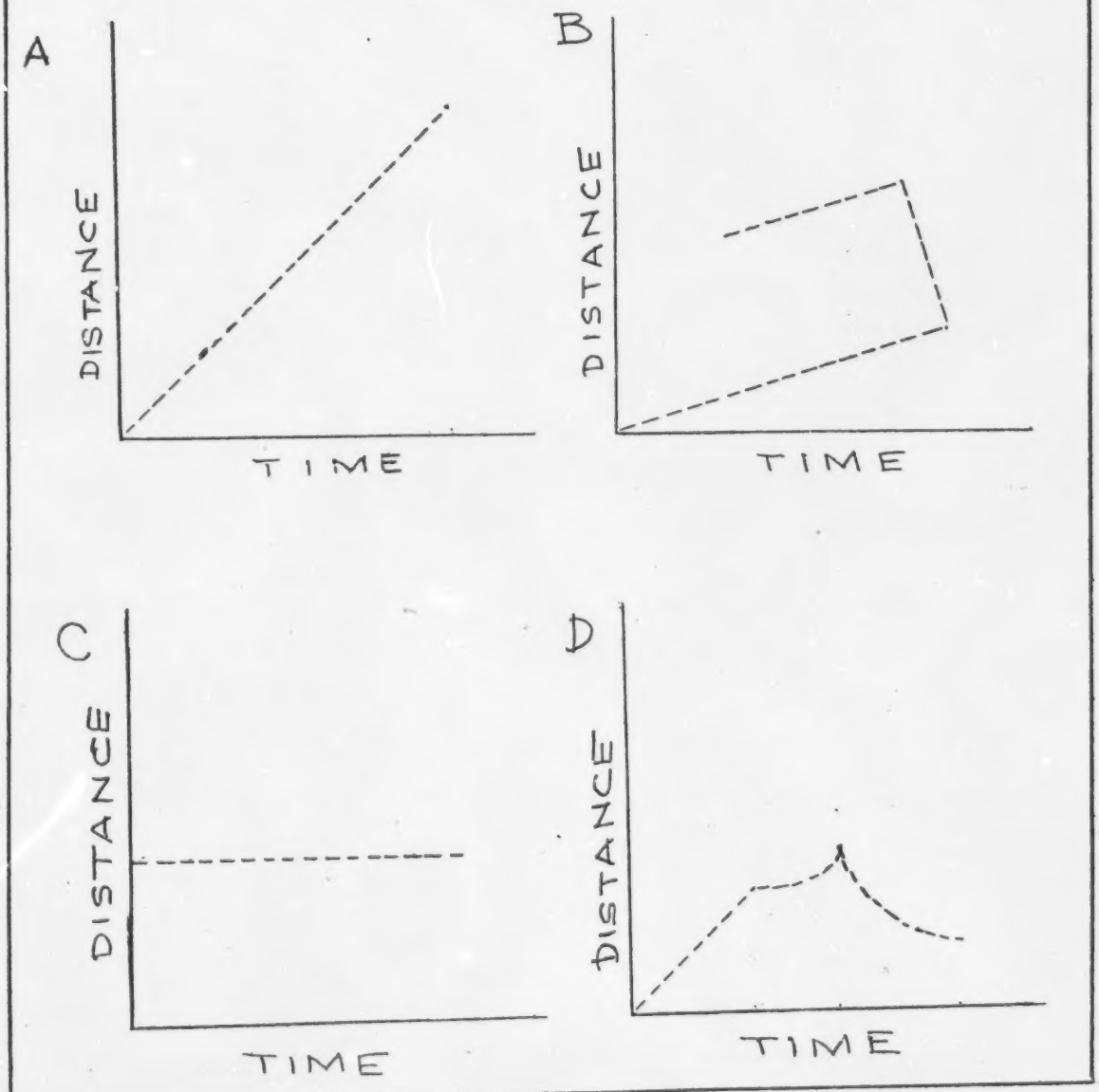
Can Read Graphs	Knows Meaning of Term: No	Knows Meaning of Term: Yes
Yes	A	D*
No	C	B

*"Correct" alternative for question

The concept being tested is "Displacement" which in this setting refers to the "distance from the starting point."

THE ITEM

A rancher has a rectangular field which measures three miles by six miles. He wishes to check the fence around the field which he can do at 3 miles per hour. If he starts at one corner and proceeds along the long side first, which of the following graphs give his DISPLACEMENT during the first five hours of this task.



The person who CHOOSES alternative "A" has identified velocity rather than displacement. In displacement, when the rancher changes direction at a corner of a field, his displacement velocity is no longer constant because it is dependent upon the distance across the field rather than the distance travelled. We may, therefore, assume that an "A" choice indicates the ability to read graphs (DO+) but not an understanding of the concept (KNOW-). In the correction of this error it is usually sufficient to point it out because the learner already has an understanding of the more generic graph reading ability.

The person who chooses alternative "B" would be in the opposite state to "A". That is, he or she has some understanding of displacement (KNOW +) but cannot read graphs (DO-). The graph represents the pathway of travel, so that the "distance from origin" aspects seems to be recognized. However, this is a TIME graph which doubles back on itself. As a result, for the duration of the upper two legs of the line on the graph the rancher would have to be in two places at the same time to satisfy the conditions of this graph. Since this event is physically impossible, the person who chose this alternative is clearly not able to read graphs.

In this case, correction of the "error" is much more difficult since merely telling the learner what he or she did wrong is usually not sufficient. Instead, a remedial program in graph reading will probably be necessary in order to correct this sort of error.

The choice of alternative C is negative in both dimensions of the problem. Since the graph shows the rancher to be a constant distance from his starting point this graph will occur in two conditions. Either he is remaining stationary or he is travelling in a circle around his starting point. Neither of these two possibilities fit the conditions specified in the problem. In either case he must have dematerialized and rematerialized at the starting time.

To correct this level of error the learner needs both the remedial graph reading program and the term explained again. In this case some of those who chose "C" will choose "A" on a second administration. These will be the learners who do not yet know the term. This possibility illustrates the common property found in my research that learners may change from a wrong answer to a different wrong answer and not exclusively from a wrong answer to a right answer as they learn.

Alternative "D" of course is the correct answer.

This item illustrates the underlying principles built into the "Test Construction of Teachers" part of the experimental program being proposed.

Any of the possible interactions between reference frame, procedure, and knowledge can be employed in this manner as can typical learner errors. The results are items which are powerfully diagnostic. In addition to this, the measure of improvement in the learner's status is not how many answers he or she answers "correctly" but which answers he or she chooses, and why.

Thus we see that with only a minor modification in the way we construct test items coupled with a major shift in the way we use the data we collect, we are now able to support a LISTENING TO approach to teaching with our system of observation.

Mathematics Example

The findings of the research I have thus far conducted suggest two important characteristics of the problem solving strategies used by learners.

First, there are several different strategies employed by people when they try to solve problems.

Second, the strategy of first choice of particular individuals tend to be stable and consistent within an individual and to vary among individuals. Of less certainty is the possibility that the order of choice may also be consistent within individuals and may also display a developmental sequence property.

It is clear that at a point of transition from one level to the next in terms of complexity of world-view that right answers peaks, followed immediately by a surge of wrong answers, in which the variety of wrong answers, the frequency of wrong answers, and the relationships among wrong answers all seem to change. This surge of wrong answers does not coincide with a valley in the sequence of right answers, but follows the peak of right answers by six months or less. Five such surges have been found thus far between the ages of 7 and 20.

A theoretical explanation for these phenomena is not yet clear. It is clear, however, that, since alternative strategies are being employed in any case, it would seem to be quite reasonable to deliberately teach and to encourage the use of alternative procedures to solve problems.

It is more typical of mathematics teachers to present the "most efficient" solution procedure for each problem type, than to encourage the exploration of alternative procedures. The proposed program would encourage the latter approach.

To illustrate this aspect of the in-service program I am proposing, I will illustrate several strategies for a Single problem class in mathematics.

Consider the problem:

EQ 1: $2x + y = 3$
EQ 2: $x + 2y = 0$
Find x and y

In the traditional approach to the teaching of mathematics, this class of problem is solved by its own algorithm. The use of the "most efficient" algorithm for each class of problem, however, may have led to many people not being very successful in learning mathematics, since there are many such algorithms, and they appear to be disconnected from each other, particularly for the "poor" student.

This attribution is very well illustrated by the typical

solution procedure I have observed to be used by those who are "unsuccessful" in mathematics.

By inspection, a possible pair of values from Equation 1 would be $x = 1$ and $y = 1$. That is, $2x + y = 3$ looks something like $2 + 1 = 3$. This "solution" is not correct in this case, but could be in some cases. This "inspection" strategy, by the way, was the one learned in elementary school to solve problems like $2 + \square = 3$.

If the habit of checking the solution in the other equation has not been thoroughly established the learner will have a solution strategy of first choice which will sometimes yield correct answers and sometimes not, for a reason unknown to the learner. Apparently, lack of success in mathematics seems to be a learned behavior.

Of course, the inspection approach will work for this problem if the learner uses the information in both equations. In Equation 2 we have a clue that one of the hidden numbers (either x or y) must be negative. We draw this conclusion because the sum is "zero" (o). By ranging either x or y through a limited number of negative values, we can very quickly come to the conclusion that the value $x = 2$ and $y = -1$ satisfies both equations. This procedure is perfectly correct; it is simply not the "most efficient" strategy for this class of problem, particularly when roots are rational numbers (fractions and mixed numbers).

There are other procedures which are perfectly valid. One is what I choose to call the "equivalence" strategy, which applies the rule "two things which are equal to the same thing are equal to each other."

Our objective is to reduce these two equations in two unknowns to a single equation. Our procedure is to reorder both equations as follows:

$$\text{New: EQI:} \quad X = \frac{3-y}{2}$$

$$\text{New: EQ2:} \quad X = -2y$$

Since the right hand values of these two equations are both equal to X, we can set them as being equal to each other:

$$\frac{3-y}{2} = -2y$$

The person who is skillful at inspection will see immediately that y = -1. If not, reordering to

$$3 - y = -4y$$

should do the trick.

The third approach is somewhat similar to the last one and is known as a "substitution" strategy. It relies upon the logical proposition that "any element in any statement can be replaced by its equivalent."

If we find a value for (x) as in New EQ2 (x = -2y) and replace (x) in Equation 1 with this value as in:

$$EQ4: \quad 2(-2y) + y = 3$$

Once again an inspection approach yields a value of (y = -1).

A fourth approach, the "imaging" approach is even more interesting. In this case, we can drop the (x's) and (y's) and operational and relational signs as follows:

	x	y	c
Row 1:	2	1	3
Row 2:	1	2	0

We now manipulate the rows by multiplication or division and combine entire rows with each other by addition or subtraction until a particular target pattern is achieved. This target or "image" is:

1	0	x
0	1	y

The procedure will look like this:

Row 1 (from above):	2	1	3	
Step 1; (Row1 X 2):	4	2	6	
Step 2; (Subt Row2):	1	2	0	(-)
	3	0	6	
Step 3; (Div by 3):	1	0	2	(New Row 1)

This last result, by making the second term in the row a zero (0), and by reducing the first term to a one (1), now conforms to the first row of the target.

We can now adjust the second row to fit the target, as follows:

Row 2 (from above):	1	2	0	
Step 1; (Subt NR1):	1	0	2	(-)
	0	2	-2	
Step 2; (Div by 2):	0	1	-1	(New Row 2)

Putting thes two end products together we get:

New Row 1:	1	0	2
New Row 2:	0	1	-1

This pattern conforms to the target image. The two right-hand numbers are the desired roots, and the four left-hand numbers form an Identity matrix.

In effect, we have used the General Linear Model matrix solution to perform this result. This latter procedure is easily programmable on a computer, and once understood, is the open door to a good deal of mathematics, particularly multivariate ststistical procedures.

Teaching strategy

Some of these approaches we will explain and demonstrate once we gain evidence that the learners will understand the approach. Others we will encourage the learners to discover for themselves by encouraging the invention of alternative approaches and the demonstration of how these approaches work by the learners themselves.

We find, in these four different strategies, useful approaches which can be used separately or in combination. More important, inspection, transforming an unknown into a known equivalent, substituting a known equivalent, or transforming a known target are all strategies which can be used outside of their mathematical applications.

The effective problem solver will be "flexible," that is, he or she will not use an habitual strategy of first choice, but will select a strategy of first choice which seems to fit the properties of the problem being solved.

Clearly, any of the roughly one dozen strategies I have identified can be explored by the procedure just described, and their efficiencies in several contexts can also be explored. The important outcome, I suspect, is that the learner develops the habit of exploring. It appears from my teaching experience that the exploratory learner tends to be able to solve a much greater range of problems than can be solved by people who rely on their habitual strategies.

But this ability to explore is not enough. To be truly effective the exploratory problem solver must also check outcomes for the purpose of detecting and correcting errors. In this respect, Argyris and Schön have made an important contribution in the interpersonal field, which, when borrowed, strengthens my work.

If we combine, in a group setting such as a classroom, exploratory behavior with public testing of assumptions and a trace-back approach when discrepencies or discontinuities are found, then a mutual learning system (Model O-II) is probably created. LISTENING TO the learner would become the operational basis for instruction, and, it appears, a major advance in learning effectiveness should be expected.

As can be seen from this discussion, the major components of this teaching strategy involve, 1) surfacing the hidden assumptions which affect outcomes; 2) observing process as well as content;

3) exploring the outcomes of alternative problem-solving strategies; and 4) checking outcomes using systematic trace-back procedures, in addition to comparison of outcomes with valid information, to find hidden errors in the reference frame, the solution strategies and/or the information base assumed. The learners would spend most of their time discovering how to detect and to correct errors.

Our ultimate objective is people who are not only self-teaching, but who can relate to others in such a way that they can develop in those others this same capacity for mutual problem solving.

The project proposed here would be expected to produce such individuals in reasonable numbers before it could be considered to be an unqualified success.

APPENDIX B

TECHNICAL NOTE

For the benefit of those who have a fairly substantial mathematical background, I will indicate briefly our current thoughts on the properties of "wrong answer" distributions as these are emerging from our analysis. I will also review briefly the implications of these findings to the emerging theory.

Dr. N. N. Shklov and I have been working for about three years, most recently with the aid of a grant from IBM, to try to develop a test theory which fits our observations of these data.

At the within item level, as illustrated on page 11, the best fitting curves for all answers (and not just WRONG ones) seems to be BETA distributions. There are also irregularities in these curves which are found in replication conditions, suggesting that the simple beta is not sufficient to describe the properties of these distributions.

For this reason, we are proposing a two level model. The "within item" model will be built upon beta distributions. The between item model will be a set of second order transforms, which will be designed to explain several properties which have been found to exist between alternatives. I will describe our approach to the within item problem first.

Since it is obvious that some form of decision is being made, when a particular alternative among several is selected, we have decided to use a decision theory model. As a result, we are trying to develop a multi-dimensional BAYSEAN decision model to describe "within item" behavior.

There are, however, several constraints we need to impose. The number of dimensions (n) of the model are the number of alternatives from which the learner must choose.

Also, if the probabilities for choosing any alternative are accumulated, their sum must be ONE (1). Therefore, if the heights

of the betas are converted to proportions of the N number of learners who wrote the test, then these must also be the Bayesean probibilities. That is

$$\sum_{i=1}^{n} \frac{\hat{y}_i}{N} = \sum_{i=1}^{n} p_i = 1.00$$

Where $\hat{y}_i$ is the frequency of the selection of the ith alternative predicted by Beta.

At the present time, this pattern does not appear to be "normal". That is, the General Linear Model does not seem to apply to these data.

The secondary level of the Beta distributions also seem to be Beta distributions. The mode of this second order Beta is at the lower end of the scale for the WRONG answers and near the upper end for the RIGHT answers. This much would be expected. However, if we plot the secondary modes to the third level, then we still get Beta distributions, with an upward shift in progressive order for the Wrong answers and downward for the Right ones.

When these four curves are aggregated for both sets of answers, there are several surprises. The height of the aggregated modal frequencies for right answers increases. The level of the bottom of the valleys remains constant. Where the wrong answers are concerned, the reverse is true. The height of the modal frequencies being about constant, with the depth of the valleys also increasing with age. Thus, increased maturity seems to be characterized both by more right answers and greater variability of answer.

Another interesting observation is that the peak frequencies of the wrong answers do not coincide with the trough frequencies for the right ones, but rather lag the peaks of the right ones by about 5 months. This observation seems to support Piaget's contention that learners need to reorder their thinking when they move from one stage to the next. There are five such transitions between the ages of 7 and 20.

When the impact of the frequencies of selection is removed by testing the n x n contingency tables for homogeneity, the joint selection of wrong alternative pairs between items far exceeds the mutual rejection of pairs, exceeds somewhat the joint selection of right answer pairs, increases in frequency with age and changes systematically with age.

It is easy to understand why the wrong answers seem to be better pridictors than right answers with all of this systematic transformation going on. And with the amount of second order transformation, it is clear that the GLM should not apply.

Transformations within items with age is expected and will be reported upon at NCME in Boston in April of 1980. There is also some evidence that strategies of first choice for a given age level may be mutually exclusive to sub-populations, perhaps reflecting some sort of "learning style."

It is expected as a result, once the second order transforms have been estimated, that the equations which result will display a set of parameters which will probably prove to have discrete rather than continuous values, and which will predict each learner's progress with a respectable level of accuracy.

Perhaps of greater overriding importance, the model should be able to indicate whether or not some pathways are more educationally productive than others. If so, we would have a measure of the effectiveness of educational interventions in terms of a shift of a learner to a more effective pathway.

As a result, should the mathematical technology be in place by the time we are summarizing the results from the projects being proposed here, definitive tests of the effectiveness of the LISTENING TO technology should result. The sample now in hand of over 4000 children aged 7 to 20, with the measures repeated after a 5 month interval, should give an excellent "control group" for the status quo or LISTENING FOR procedure.

Other questions may also be open for exploration. For instance,

it is probable the "exploratory learning" technology developed by myself, using wrong answers, and the "Double-Loop" learning technology, developed by Argyris and Schön are complementary rather than equivalent. If this approach to data analysis proves to be a sensitive as present indications suggest to differential treatments, then this sort of question would be easy to answer.

Still another interesting direction is the finding that personality variables seem to be more highly related to wrong answers than to right answers. In this case, the linkage between the worldview of the person, his or her ethical standards, and his or her learning effectiveness may be open to access.

Another direction which might be openned relates to the nature of effective teaching. Not just in general terms, but which approaches are most effective for which types of learner?

A further technical consideration involves the development of refinements in the statistical procedures now under study, and in the appropriate methods for test construction to make the best uses of these sources of data. In this way we should be able to take full advantage of the information available within the alternative technology.

Finally, once this latter step is in progress, ways need to be found of reducing the complexity of these processes, perhaps with the aid of electronic data processing, so that the improved information base from this fine tuning will be readily available to teachers in the field. It would be of no great benefit if teachers had to become professional logicians or professional statisticians in order to conduct their day-to-day decision making in the classroom. This last extrapolation is a fascinating and challenging prospect indeed.

APPENDIX C

Bibliography

This bibliography lists the major sources for the information presented in this "Prospectus" and indicates availability. In the case of my papers, I have also indicated which aspects of the evidence are contained in each paper.

Argyris, Chris, and Schön, Donald A., (1974) Theory in Practice. San Fransisco, Jossey-Bass.

Argyris, Chris, and Schön, Donald A., (1978) Organizational Learning; A Theory of Action Perspective, Reading, Mass., Addison-Wesley.

Bateson, Gregory, (1958) Naven, Stanford Cal., Stanford University Press. Bateson derives the term Deutero-learning from Ashby, Ross, (1954) Design for a Brain. London, Chapman Hall.

Piaget, Jean, (1976) The Essential Piaget. Edited by Howard Gruber and Jacques Voneche, New York, Basic Books.

Powell, J. C.,(1968) The interpretation of wrong answers from a multiple choice test. Educational and Psychological Measurement, 28(2), 408-412. This paper reports the results from my Qualifying Project.

Powell, J. C., (1970) Achievement Information from Wrong Answers. (Short Title). Unpublished Ph. D. Dissertation, University of Alberta, Edmonton, Alberta. First in-depth exploration which established many of the findings replicated in my subsequent work.

Powell, J. C., (1976) Evidence for a phase and stage developmental sequence derived from response patterns on multiple choice tests. Paper presented to the American Psychological Association, (62pp). Available from ERIC. This report summarizes all of the earlier studies, including the one listed next, and gives several findings not available elsewhere.

Powell, J. C., (1977) The developmental sequence of cognition as revealed by wrong answers. Alberta Journal of Educational Research, 23(1), 43-51. Reports the wrong answer reasoning of 550 children from 8 to 14+ years of age. The finding of a strong developmental sequence using a different test and age group from the one in my Dissertation supports the finding of a wrong answer hierarchy among adults in the earlier study.

Powell, J. C., (1978a) Cognitive development from wrong answers. Paper presented to the International Association for Applied Psychology, Munich, West Germany. Available from author. Reports the discovery of a strong relationship between wrong answers and personality variables.

Powell, J. C. , (1978b) Wrong Answers on multiple choice tests; Blind guesses or systematic choices? Paper presented to the Psychometric Society, Hamilton, Ontario. Available from author. A confirmatory study using Bock's MULTIQUAL that answers patterns are non-linear.

Powell, J. C., (1979), Can developmental status information be obtained from wrong answers? Papar presented to the National Council for Measurement in Education. San Fransisco. Available from ERIC. This paper reports the results of an exploration for relationships among pairs of wrong answers when the impact of frequencies is removed. Developmental patterns favoring wrong answers over right answers were found.

Powell, J. C., (1980) (May have a co-author) Patterns of change among answer selection. Paper to be presented to the National Council for Measurement in Education, Boston Mass. This paper will report the changes in answer selection across a 5 month interval among more than 3000 children and youth from 7 to 20 years of age. Relationships with the impact of frequencies removed will be sought and explored for their distribution properties.

Powell, J. C., and Isbister, Alvin G., (1974) A comparison between right and wrong answers on a multiple choice test. Educational and Psychological Measurement, 34(3), 499-509. This paper presents the details of one of the validation studies on the special test used in my Dissertation. It gives the details of the construction of this test, and explores several aspects of the "Know-Guess" hypothesis. The findings refute this hypothesis on all of the aspects examined.